CIRJE-F- 1234

Probability-based A/B testing with Adaptive Minimax Regret (AMR) Criterion for Long-Term Customer Metrics

Makoto Abe
The University of Tokyo

August 2024

Probability-based A/B testing with Adaptive Minimax Regret (AMR) criterion for long-term customer metrics

Makoto Abe

Graduate School of Economics
The University of Tokyo
Email: abe@e.u-tokyo.ac.jp

August 20, 2024

Probability-based A/B testing with Adaptive Minimax Regret (AMR) criterion for long-term customer metrics

ABSTRACT

In economics, uncertainty is distinguished into two types: risk, which can be evaluated in terms of probability, and ambiguity, in which the probability is unknown. In decision making under risk, the rational course of action is to make a choice that maximizes expected utility, which is the utility of an event weighted by its probability. On the other hand, under ambiguity, where the probability is unknown, how should decisions be made?

We first introduce the Minimax Regret, a decision-making criterion under ambiguity where probabilities are unknown but the interval is known. As a concrete example, consider two slot machines: one existing and one new. The winning probability of the former is known, while the winning probability of the latter is unknown, with only the interval provided. In this case, the optimal strategy according to the Minimax Regret criterion would be to randomly pull each of the two slot machines with a certain probability.

Next, when utility is measured by a long-term metric, the interval of uncertainty for this metric decreases over time. To address this, we introduce the Adaptive Minimax Regret (AMR) approach, which maximizes utility by updating the probabilities according to the Minimax Regret criterion based on the information available at each point in time. Simulation testing on the case of the existing and new slot machines mentioned earlier showed that AMR produced high performance comparable to bandit algorithms. As an application of AMR in marketing, we propose sequential campaign strategies and probabilistic A/B testing aimed at maximizing the average customer lifetime (utility) of the target audience.

Keywords: Minimax Regret, Bandit Algorithm, Probabilistic A/B Testing

1. CONCEPTUAL FRAMEWORK

As the importance of long-term customer metrics such as lifetime value and churn rate continues to grow, companies find themselves compelled to make prompt decisions before observing the full results.

First, let us consider the following scenario:

In an existing campaign, E, with a track record of 4 years, the average lifetime of acquired customers was 2.7 years. Now, to increase the lifetime, a new campaign, N, is being planned. If we were to implement N and compare its effectiveness with E, we need to wait for 4 years for the results to come out. If, however, we knew the probability distribution of the average lifetime of N from prior market research (e.g., 2 years with 0.4 and 4 years with 0.6), we would choose N with the higher expected value (3.2 years) over E (2.7 years). However, under "ambiguity," where this probability distribution is unknown but only its interval (support) is known, how should decisions be made?

This situation is related to decision-making in situations where uncertainty is distinguished into two types: 'risk', which can be assessed with probability, and 'ambiguity', where the probability is unknown (Savage 1951). In decision-making under "risk," the Bayesian decision is rational, where one chooses the option with the higher expected utility weighted by the probabilities of outcomes. On the other hand, in decision-making under "ambiguity," where probabilities are unknown and only intervals are known, criteria such as the Maximin (maximizing the minimum utility), Minimax Regret (minimizing the maximum regret), and Maximax (maximizing the maximum utility) have been proposed.

Here, we propose an approach called Adaptive Minimax Regret (AMR), applying the Minimax Regret criterion to the selection of campaigns E and N based on the results obtained at each point in time and updating it sequentially (Manski 2011).

2. RELEVANT LITERATURE

The minimax regret criterion is a decision-making principle that aims to minimize the maximum possible loss in utility. On the other hand, the maximin criterion tends to be overly conservative in decision-making as it focuses on maximizing the worst-case scenario. Conversely, the maximax criterion tends to be overly optimistic as it aims to maximize the best-case scenario. The key feature of the minimax regret criterion is that it strikes a balance in decision-making orientation (Manski 2007).

The second feature is that decisions are made probabilistically. In the previous campaign example, Minimax Regret is achieved by randomly implementing E and N with probabilities 1-d and d, respectively, as shown in the table.

campaign	**utility** **$u(\cdot)$**	**implementation probability**
E: existing	a fixed value, known	1-d
N: new	$b \in [\, b_L, b_U \,]$ pdf is unknown	d

Regret *R* is defined as the difference between the utility obtained from the best action ($\max(a, b)$) and the utility obtained from the chosen action ($u(d,b)$), as in (1).

(1) $$R(d, b) = \max(a, b) - u(d, b) \quad \text{where } d = \text{probability of choosing N}$$

Minimax Regret criterion minimizes the maximum regret (worst scenario) as in (2).

(2) $$\min_{d} \max_{b}\{R(d, b)\} = \min_{d} \max_{b}\{\max(a, b) - u(d, b)\}$$

The optimal solution d^* is represented by (3) (see Appendix 1).

(3) $$d^* = \frac{b_U - a}{b_U - b_L}$$

Expanding the situation to where the utility a for E is also unknown to the lesser extent and only its interval (a_L, a_U) is known, the optimal d^* is represented by (4) (see Appendix 2).

(4) $$d^* = \frac{b_U - a_L}{a_U - a_L + b_U - b_L} \quad \text{where } b_L < a_L < a_H < b_H$$

3. METHOD

3.1. Adaptive Minimax Regret (AMR)

Uncertainty in a long-term metric (utility), defined by its bounds [b_L, b_U], decreases as new data become available over time. Therefore, AMR is an approach that updates the optimal probability d^* of Minimax Regret based on the information obtained up to that point in time.

In our campaign case, $b_L(t)$ and $b_U(t)$ are updated based on the observed customers' survival up to period t. Based on the optimal probability d_t^* derived from Equation (3) at that point in time, E and N are selected and implemented.

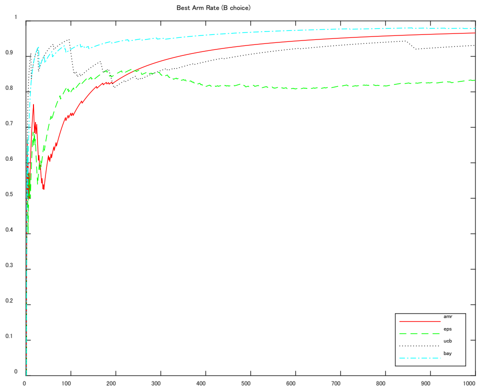
3.2. Comparison with Bandit Algorithms

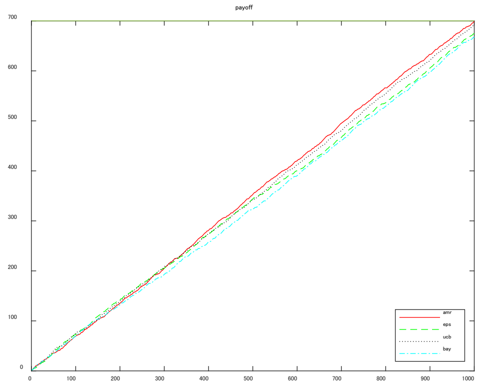
AMR shares similarities with the framework of bandit algorithms, which aim to maximize cumulative return by continuously pulling the arm with the higher return, given the uncertainty about which arm (E or N) yields higher returns. In bandit algorithms, balancing "exploitation" (pulling the arm with a higher return) and "exploration" (inferring the return of each arm) is crucial. Various methods have been proposed for this purpose (Lattimore and Szepesvári 2020). Therefore, the effectiveness of AMR is first compared with three commonly used bandit algorithm methods (ε-greedy, Upper Confidence Bound, Thompson Sampling).

4. SIMULATION STUDY

In this study, two arms A and B with different winning probabilities (a=0.5, b=0.7) are prepared. The objective is to maximize the cumulative number of wins while continuously pulling the arms 1000 times, with unknown winning probabilities for both arms. After each pull, the result of a win/loss is observed. In AMR, the uncertainty in the winning probability of each arm is updated based on the information obtained up to that point, using the 95% confidence interval estimated from the data.

The left figure plots the probability of pulling the arm with the higher return (B) against the number of pulls. Initially, exploration is conducted through trial and error, but eventually, the probability of pulling B approaches 1. AMR performs second best after Thompson Sampling based on Bayesian methods. The right figure plots the cumulative return against the number of pulls. If only arm B is continuously pulled, the line would have a slope of 0.7. In this study, AMR showed the best performance.





This simulation experiment showed that AMR produces performance that is comparable to or exceeds that of existing methods commonly used in bandit algorithms.

5. MARKETING APPLICATION

The bandit algorithm cleverly exploits the improvement in the estimation accuracy of the static parameter, which is the winning probability of the arms, as the number of trials increases. The true advantage of AMR lies in its ability to leverage dynamic parameters in situations where they are censored, which cannot be addressed by traditional bandit algorithms. Let us examine this in the context of the previous campaign case, focusing on customer lifetime.

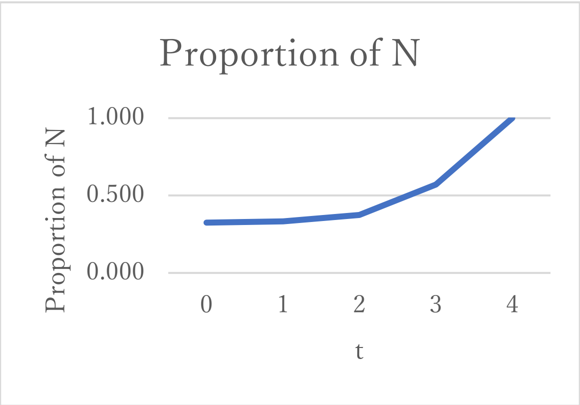
In the existing campaign E, the customer retention rates for all periods t are known, and the average lifetime is calculated to be 2.7 years (=0.8+0.7+0.6+0.6). A new campaign N is under consideration, and the goal is to utilize strategies in E and N to increase the average lifetime during the first 4 years (ignoring 5^{th} year and beyond).

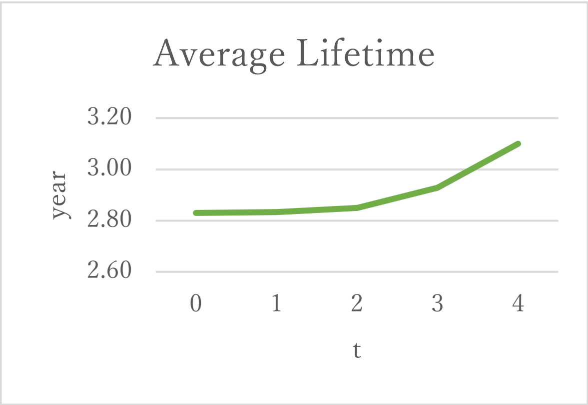
Since there is no customer retention rate data for N, the average lifetime is unknown until it is actually implemented. After implementation, the customer retention rate for each period *t* (0.9, 0.8, 0.7, 0.7) is observed, reducing the uncertainty in the average lifetime ($b_L(t)$ and $b_U(t)$). Based on the survival information obtained up to period t, AMR updates the optimal mixing probability d_t*. The results are summarized in the table below.

year	survival probability		lifetime bound		proportion of N	ave. lifetime
t	E	N	b_L	b_H	d_t*	
0	1	1	0	4	0.325	2.83
1	0.8	0.9	0.9	3.6	0.333	2.83
2	0.7	0.8	1.7	3.3	0.375	2.85
3	0.6	0.7	2.4	3.1	0.571	2.93
4	0.6	0.7	3.1	3.1	1.000	3.1
ave. lifetime	2.7	3.1				3.1

For example, in row t = 2, the probability of survival of 0 years (less than 1 year) is 0.1 (=1-0.9) and the probability of survival of 1 year (between 1 and 2 years) is 0.1 (=0.9-0.8). Therefore with the remaining probability of 0.8, the shortest survival is 2 years and the longest is 4 years. This gives us b_L=1.7 (=0.1x0+0.1x1+0.8x2) and b_H=3.3 (=0.1x0+0.1x1+0.8x4).

The average lifetime for N is 3.1 years, but this information is not known until year 4. As time progresses, the interval of the average lifetime decreases. Correspondingly, the optimal ratio d_t* for N, determined by the Minimax Regret criterion, increases and eventually reaches 1 (left figure). By implementing a campaign that mixes E and N with probability d_t*, the average lifetime extends from 2.83 years to 3.1 years (right figure).





6. CONCLUSIONS

In this study, we proposed the AMR approach, which mixes the selection of campaigns E and N optimally with a certain probability. AMR applies the Minimax Regret criterion based on the result of N obtained at each point in time, updating the mixture probability sequentially.

Evaluating long-term metrics like customer lifetime requires an extended observation period. Even with big data in the absence of statistical error, these metrics cannot be fully captured. Companies can calculate the optimal mixing probability based on current information at hand and update it sequentially, enabling them to respond before knowing the complete results of long-term metrics. As a long-term metric in marketing, AMR can also be applied to areas beyond customer lifetime, such as the repeat rate of products, subscriptions, and services.

REFERENCES

Lattimore, T. and C. Szepesvári (2020), Bandit Algorithms, Cambridge University Press.

Manski, C. (2007), Identification for Prediction and Decision, Cambridge, Mass.: Harvard University Press.

Manski, C. (2011), "Choosing Treatment Policies Under Ambiguity," Annual Review of Economics, 2011.3: 25–49.

Savage, L. J. (1954), The foundations of statistics, John Wiley & Sons.

Appendix 1: Derivation of *d** for Minimax Regret Criterion when *b* is unknown

Recall the payoff of the two campaigns E and N.

campaign	utility $u(\cdot)$	implementation probability
E: existing	a fixed value, known	1-d
N: new	$b \in [\, b_L, b_U \,]$ pdf is unknown	d

Without loss of generality, assume that the support (uncertainty interval) of b is such that $b_L < a < b_H$.

Utility $u(\cdot)$ is express as a function of d and the unknown payoff b as (A1).

(A1) $u(d,b) = (1-d)a + d \cdot b = a + (b-a)d$

Since the regret is defined as follows,

$$R(d,b) = \max(a,b) - u(d,b) \quad \text{where } b \text{ is unknown}$$

decision d under the Minimax regret criterion is shown in (A2)

(A2) $\min_d \max_b \{R(d,b)\} = \min_d \max_b \{\max(a,b) - u(d,b)\}$

Maximum regret $\max_b \{R(d,b)\}$ is expressed as (A3).

$$\begin{aligned}
&\max_b \{\max(a,b) - u(d,b)\} \\
&\qquad = \max_b \{l(b>a)[b - \{a + (b-a)d\}] + l(a>b)[a - \{a + (b-a)d\}]\} \\
&\qquad = \max_b \{l(b>a)(b-a)(1-d) + l(a>b)(a-b)d\} \\
&\qquad = \max\{(b_U - a)(1-d), (a-b_L)d\}
\end{aligned}$$

(A3)

Since decision d^* that miminizes (A3) satisfies $(b_U - a)(1-d^*) = (a - b_L)d^*$,

(3) $d^* = \frac{b_U - a}{b_U - b_L}$

Figure A1 depicts the maximum regret as a function of d and the optimum decision d^*.

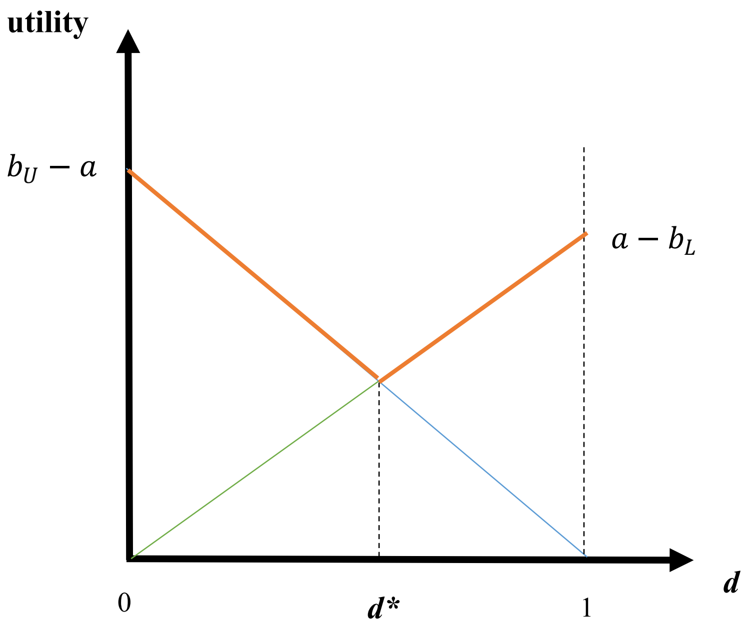


Figure A1. ***d**** **for minimax regret criterion when** ***b*** **is unknown.**

Appendix 2: Derivation of *d** when both *a* and *b* are unknown

Without loss of generality, assume that the supports (uncertainty intervals) of *a* and *b* is such that $b_L < a_L < a_H < b_H$ (*b* is more uncertain).

Utility $u(\cdot)$ is express as a function of *d* and the unknown payoffs *a* and *b* as (A4).

(A4) $$u(d, b, a) = (1 - d)a + d \cdot b = a + (b - a)d$$

Maximum regret $\max_b\{R(d, b, a)\}$ is expressed as (A5).

$$\max_{b,a}\{\max(a, b) - u(d, b, a)\}$$
$$= \max_{b,a}\{l(b > a)[b - \{a + (b - a)d\}] + l(a > b)[a - \{a + (b - a)d\}]\}$$
$$= \max_{b,a}\{l(b > a)(b - a)(1 - d) + l(a > b)(a - b)d\}$$
$$= \max\{(b_U - a_L)(1 - d), (a_U - b_L)d\}$$

(A5)

Since decision *d** that miminizes (A3) satisfies $(b_U - a_L)(1 - d^*) = (a_U - b_L)d^*$

(4) $$d^* = \frac{b_U - a_L}{a_U - a_L + b_U - b_L}$$

Figure A2 depicts the maximum regret as a function of *d* and the optimum decision d^*.

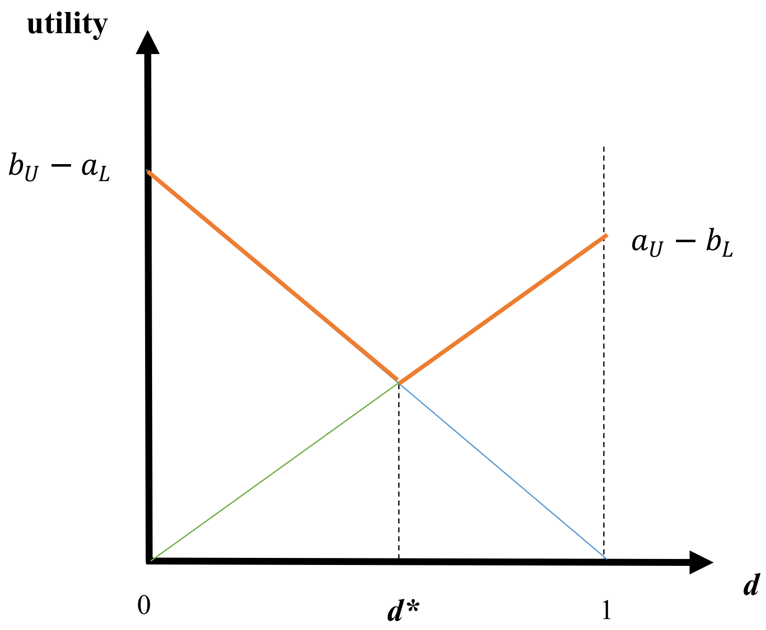


Figure A2. *d for minimax regret criterion when both *a* and *b* is unknown.**